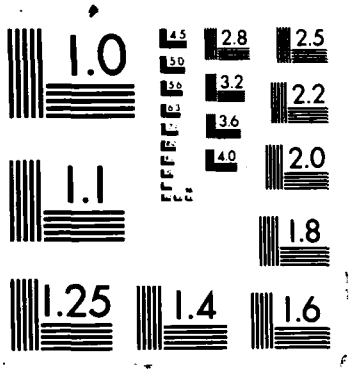
AD-A181 103 TOWARD AN ARTIFICIAL INTELLIGENCE ENVIRONMENT FOR DTIC (DEFENSE TECHNICAL (U) DEFENSE TECHNICAL INFORMATION CENTER ALEXANDRIA VA OFFICE OF I A D KUHN MAY 87 1/1

UNCLASSIFIED DTIC/TR-87/13 F/G 5/2 NL


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3

DTIC/TR-87/13

AD-A181 103

AD-A181 103

TOWARD AN ARTIFICIAL INTELLIGENCE ENVIRONMENT FOR DTIC: PROPOSED TASKS, RECOMMENDED CONFIGURATIONS, PROJECTED START-UP COSTS

Allan D. Kuhn

May 1987

DTIC Defense Technical Information Center

DTIC AI Foundational Series No. 3

Office of Information Systems and Technology

Cameron Station, Alexandria, VA 22304-6145

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified/Unlimited	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution ulimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) DTIC/TR-87/13	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Defense Technical Information Center 6b. OFFICE SYMBOL (If applicable) DTIC	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Cameron Station Alexandria, VA 22304-6145	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 65801S; PROJECT NO.; TASK NO.; WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Toward an Artificial Intelligence Environment for DTIC: Proposed Tasks; Recommended Configurations; Projected Start-Up Costs.

12. PERSONAL AUTHOR(S)
Allan D. Kuhn

13a. TYPE OF REPORT	13b. TIME COVERED FROM ____ TO ____	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
		8705	15

16. SUPPLEMENTARY NOTATION
DTIC AI Foundational Series No. 3

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	
5	2		Management, Artificial intelligence, AI, Defense Technical Information Center. DTIC, DoD Gateway Information System, DGIS, Information technology,
5	1		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report discusses candidate Artificial Intelligence (AI) tasks for initiating AI projects at DTIC, several configuration issues, and start-up costs of expertise and equipment. Configuration issues concern Mission, organization, and system. The DTIC AI mission wil focus on prototyping practical applications of high value to the DTIC user community. DTIC has the choice of establishing a wholly self-contained AI activity, or participating in, contributing to, and sharing the resources of the Defense Applied Information Technology Center (DAITC). Coordinating AI functionality will be necessary for incorporating AI development activities and implementations on the DoD Gateway Information System (DGIS), the core computer system fo the DAITC. Workstation consultant mode and backend storage are two options for production programs management. Two cost projection models are given as start-up costs for an AI environment. The costs are for expertise and hardware/software. AI expertise and AI ware are mutually dependent for a strong AI environment.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	Unclassified/Unlimited
22a. NAME OF RESPONSIBLE INDIVIDUAL Allan D. Kuhn	22b. TELEPHONE (Include Area Code) (202) 274-5367
	22c. OFFICE SYMBOL DTIC-EB

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

>>>

TOWARD AN ARTIFICIAL INTELLIGENCE ENVIRONMENT FOR DTIC: PROPOSED TASKS; RECOMMENDED CONFIGURATIONS; PROJECTED START-UP COSTS

Allan D. Kuhn

May 1987

DTIC AI Foundational Series No. 3

Defense Technical Information Center
Office of Information Systems and Technology
Alexandria, VA 22304-6145

>>>

DTIC AI FOUNDATIONAL SERIES

No. 1: **Toward An Artificial Intelligence Environment for DTIC: Staffing Qualification Criteria For AI Application Development. Defense Technical Information Center, Feb 87, AD-A181 100.**

No. 2: **Artificial Intelligence Developments Re: DoD Gateway Information System (DGIS) & Defense Applied Information Technology Center (DAITC). Defense Technical Information Center, Feb 87, AD-A181 101.**

No. 3: **Toward An Artificial Intelligence Environment for DTIC: Proposed Tasks; Recommended Configurations; Projected Start-up Costs. Defense Technical Information Center, May 87, AD-A181 103.**

No. 4: **[Pending]**

TOWARD AN ARTIFICIAL INTELLIGENCE ENVIRONMENT FOR DTIC:
PROPOSED TASKS; RECOMMENDED CONFIGURATIONS; PROJECTED START-UP COSTS

EXECUTIVE SUMMARY

This paper is a companion paper to Series No. 1, Staffing Qualification Criteria for AI Application Development. This paper discusses candidate AI tasks for initiating AI projects, several configuration issues, and start-up costs of expertise and equipment for initiating AI utilization at DTIC.

Proposed tasks include:

Common Command Language (CCL) (AI-based expansion)
Diverse Database Query Expert System
Thesauri Integration for Expert Searching
Information Processing Systems
Numeric Information Query and Processing System

Configuration issues concern Mission, organization, and system.

The DTIC AI mission will focus on prototyping practical applications of high value to the DTIC user community. In order to achieve a DTIC AI environment in which applications development and prototyping will excel, expertise and equipment are needed.

DTIC at this point has the choice of either setting up and wholly supporting a self-contained AI activity, or participating in, contributing to, and sharing the resources of the Defense Applied Information Technology Center (DAITC). The DAITC, of which DTIC is a sponsor and for which DTIC is to provide the Federal Government management, has already established an embryonic AI laboratory. Contributing to this resource and taking full advantage of it while assuming its management would be both economical and cost-effective.

Coordinating AI functionality will be necessary for incorporating AI development activities and implementations on the DoD Gateway Information System (DGIS), the core computer system of the DAITC. System possibilities are workstation consultant mode and backend programs storage, with a recommendation for the latter for better production programs management on the system.

Two cost projection models are given as start-up costs for an AI environment. The costs are for expertise and hardware/software. Though the costs are seemingly high, AI expertise and AI ware are mutually dependent for a strong product orientation. But the benefits will be immeasurable in overcoming the barriers of user-machine communication, and make the human-machine interface more human-like in its functions and responses.

CONTENT

TOWARD AN ARTIFICIAL INTELLIGENCE ENVIRONMENT FOR DTIC: PROPOSED TASKS; RECOMMENDED CONFIGURATIONS; PROJECTED START-UP COSTS

Allan D. Kuhn
Defense Technical Information Center

I. INTRODUCTION

In projecting and recommending the establishment of an artificial intelligence (AI) environment for the Defense Technical Information Center, two fundamental elements are to be considered. One element is the organization of the activity, the other is the staffing of the activity. The organization concerns identifying the purpose of and needs for an AI effort. The staffing concerns the expertise for the effort.

The qualification criteria for staffing are described in DTIC AI Foundational Series No. 1 [1]. Although it seems obvious that staffing criteria cannot be established without first identifying the purpose of a DTIC AI activity, that paper also briefly reviews the general developments that have taken place in the AI community, technical problem areas experienced by the community, and current AI effort areas. Looking at these areas gave us a general view of AI and highlights development opportunities germane to DTIC interests. The elements of a DTIC effort were also initially identified, in order to refine criteria for staffing.

A major element to be considered for incorporating AI tools is the configuration of our base system, the DoD Gateway Information System (DGIS). The DGIS, which is the core computer system of the Defense Applied Information Technology Center (DAITC), serves two basic purposes. The first, the mission of the DGIS, is as a service to the DoD community for accessing the plethora of information systems, aggregating information from them, and processing the aggregated information into a product that is useful to the user. This mission is described in two DTIC reports [2][3] that explain the concepts of using DGIS. The second purpose of DGIS is as a prototype environment for enhancing information access for the DoD community. The uses and configuration of the system are mutually influential, with needs determining configuration, but configuration determining applications.

This paper, No. 3 of the DTIC AI Foundational Series, explores AI development strategies relevant to DGIS. This paper explores proposed AI development tasks, the AI configuration needed to support these tasks, and projected start-up costs to create the required AI environment.

II. PROPOSED TASKS

This section describes effort which have been identified as candidates for application of AI technology. The main criterion for identifying such efforts is the beneficial effect on and for the user. In Report Series No. 2 [4], it was point out that the application of both AI and AI-like technology has the effect of making the electronic system appear to have human-like responses. These "human-like" responses make it easier for DoD managers to successfully interact with and extract relevant data from information systems. DTIC's first excursion into the AI arena involved the acquisition of PROLOG for the DGIS Common Command Language (CCL) project.

DTIC has experienced first-hand the dendritic effect of venturing into AI: requirements and applications will generate as developments progress. For example, in the case of the CCL project, the incorporation of PROLOG immediately allows the the creation of a command language knowledge base system (KBS) as a searcher assistant. KBS development, however, leads to the need for an expert system (ES) building tool; such a tool not only represents higher level AI application, but also makes use of a developed application that is recognized as current technology. The inclusion of KBS/ES technologies generate further highly useful applications, which in the example of CCL is a CCL user profile KBS. This KBS will permit the user to tailor a CCL query system to meet one's particular needs. Not the least is the eventual employment of natural language in eliciting and formulating the user's query.

The very basic goal in implementing AI applications, then, is to make the human-machine interface more human-like, as indicated by the following tasks.

1. COMMON COMMAND LANGUAGE

There are two primary elements that establish the need for an AI-based common command language (CCL) on the DGIS. The first is to be able to "talk" with a multitude of databases without needing to know the different native command languages for them. The second is being able to address databases without needing to know their individual search mechanistics. An algorithmic program, e.g., C, can be written to create a CCL capability as a simple substitute for native command languages. The varying mechanistics, however, have generated the potential of incorporating AI capabilities. An AI-based CCL, working out of knowledge base systems for command languages and user profiles, can create a human-machine interface that is more human in its processes.

2. DIVERSE DATABASE QUERY EXPERT SYSTEM

The refinement of the query ("Is this what I am really asking?"), and transmission of the query to invoke the search relative to the database functionality. This system would include the following subsystems:

a. Query Analysis System -

System for refining the user's search query, patterned after standard reference librarian query refinement models, in natural language.

b. Subject Searching System -

Extension of the Query Analysis System, making use of multiple database thesauri.

c. Full Text Search System -

Inclusion of text search applications as a component; it would be used for retrieving not only full text stored documents, but also on all text data, e.g., the titles and abstracts of bibliographic databases. Searching documents in their language implies the need for making use of the the user's language, i.e., natural language querying; this in turn permits greater specificity, which results with increased topic relevance and context. The system needs to provide for retrieving user-designated size sections containing the user's topic in context, for browsing documents online, and capabilities for analysing and processing text information.

d. Information Scan System -

The scanning of user-selected sets of databases for query results, to determine the results volume and consequently the usefulness of each database in the set pertinent to the query.

e. Trending and Projecting System -

Statistical-based trending, along the lines of management trending and projecting packages, of user-selected databases, including in sets. For example, comparison trending and projecting would be made on two or more terms or topics.

3. THESAURI INTEGRATION FOR EXPERT SEARCHING

Development of a system to apply external database thesauri for determining relevant databases in response to a user's query.

4. INFORMATION PROCESSOR SYSTEMS

After information has been aggregated from diverse sources, the user needs the capability to standardize, process and reformat that aggregation into a product that is most useful to him. DGIS already has these capabilities to a certain extent, programmed in C. AI-based systems that would enhance and expand the quality of DGIS information processing would include a duplicate and irrelevant data analysis system, a relevant data analysis system, analysis graphics, and a tailored product system. These systems would comprise the whole of user-invoked DGIS information processing as follows:

a. Translator Generator -

An expert system for both programmer and user creation of a generator for translating an external database record format hitherto not existent in DGIS, into the DGIS standard format. Initially, this generator might be constituted of a selection of primary fields, based on the information displayed in the final product.

b. Duplicate and Irrelevant Data Analysis System -

Realizing that aggregation of information on a selected subject from diverse but pertinent information systems can result in redundant citations, this system would not only recognize redundancies, but also allow the user to determine which citations take precedence for remaining in the aggregation. This system should also recognize the lesser occurrence of irrelevant citations based on subject disjuncture with the searcher's original query.

c. Relevant Data Analysis System -

The system for doing analysis routines on the aggregated record query relevances, invoked at the discretion of the user. These routines, based on those currently available in DGIS, would perform statistical and text analyses that are bibliometrically useful to the user. AI-based routines would expand and deepen the functions of current C-based routines.

d. Information Analysis Graphics System -

In that a picture to the human mind is worth a thousand words, analytical graphics capabilities should be incorporated. ASCII-based graphics representation should also be included for those users not having bit-map terminals. Examples of statistical graphics are: occurrence of records by year, author, subject term; graphics comparison of occurrences; bar, pie, and line charts; et al.

e. Tailored Product Generation System -

This system would tailor presentable information products, in formats ascertained by the enduser. Such a system might be considered a desk-top publishing system for 1-n copies of a document, to include the standard components of a document such as cover, content, tables, graphics, and indices.

5. ROUTINE GENERATOR EXPERT SYSTEM

This is an expert system for creating generators for routine processes. This system would be used by both programmers and users, as appropriate. The expert system must make use of bidirectional natural language to define the routine, state the requirements, and build the generator. Examples of identified routines for which generators might be created are:

o Connection Agent Generator - To establish connects to remote databases not yet in DGIS, but needed by a particular user on a recurring basis. If the connect would not be incorporated in the DGIS connection routines, it could at least be retained by the user.

o Translator Generator for Common Command Language - To create the translation of a remote database command language that has not yet been done for DGIS, but is needed by a particular user. A user-generated translation could initially be based on a set of basic commands most commonly used. The ability to toggle back and forth between CCL and the native command language would be automatically included.

o Translator Generator for Information Processing - Discussed above in section on Information Processing Systems.

6. FOREIGN HUMAN LANGUAGE INTERFACES

The incorporation of human language programs that provide passive and interactive machine translation and transmission, as follows:

a. Translation of foreign language information retrieved from foreign databases.

b. In-process translation into English of messages received on the electronic mail from foreign language speaking people, with reciprocal translation from English.

c. Communicating interactively with foreign language databases, with interactive translation between English and the language of the database or record.

d. Interactive translated communication with foreign language speaking people while making use of the DGIS capability to "talk" interactively via the keyboard. An extension would be translation of voice-generated data.

7. NUMERIC INFORMATION QUERY AND PROCESSING SYSTEM

A system for identifying numeric information sources, and aggregating, analyzing, and synthesizing that information.

8. INFORMATION ANALYSIS AND EXPERT SYSTEM APPLICATION SYSTEM

A system that analyzes the information content of a targetted universe, clarifies the purpose of the intended Expert System, and recommends the standard categorical Expert System functionality. This system would be used in the building of expert systems as needed for the DGIS, plus expert system developments made for application outside DGIS.

9. PORTABLE NATURAL LANGUAGE INTERFACE

Natural language interface commonly usable with programs, routines, functions, and other areas and aspects of the electronic system.

10. NATURAL ENGLISH LANGUAGE INTERFACE TO UNIX

Natural English language interface to cryptic UNIX, to include UNIX function identification by the user's natural language description.

III. RECOMMENDED CONFIGURATIONS

The DTIC AI environment is by no means to be considered an AI research activity but an AI applications activity. Our purpose is to take both developed and leading edge applications for enhancing information accessing for our users. The following discusses configurations for mission, organization, and system.

1. MISSION CONFIGURATION

It is recommended that the DTIC AI activity mission be:

a. Identify DoD information problem areas,

b. Evaluate those problem areas for potential AI applications,

c. Build applications using currently available ware while at the same time using leading edge technology,

d. Implement those applications in a timely manner.

The DTIC AI applications development will have a product orientation. The quality and reliability of the activity will be based on its products. It is recommended, therefore, that the first activity be to construct a relatively easy-to-build system as rapidly as possible for the purpose of not only having a product, but also a demonstrable system that shows the usefulness of AI applications. This initial system would also serve as a learning exercise in product building.

The AI activity should be staffed based on a set of high quality criteria in order to excel in its mission. The activity should also form liaisons with other DoD and government AI activities, for awareness and cooperation on AI development and implementation, especially in information.

2. ORGANIZATION CONFIGURATIONS

DTIC has two options currently available for an organization configuration:

a. DTIC-sponsored and monitored AI development activity.

DTIC has already begun AI activity on DGIS through the Common Command Language effort. This development opened the door to AI for DTIC, beginning with the acquisition of PROLOG for installation on the DGIS. The potential for expanding AI development is limited primarily by funding constraints.

b. DTIC participant in the DAITC AI laboratory.

The Defense Applied Information Technology Center (DAITC), of which DTIC is a sponsor, and for which DTIC is to provide the Federal Government management, has already established an embryonic AI Laboratory. Contributing to this resource and taking full advantage of it would be both economical and cost effective. It is planned that eventually DTIC will provide the Federal manager for this Laboratory. Once that is achieved, the AI activities of the DAITC Laboratory and of DGIS could be merged and coordinated to improve cost efficiencies and accelerate progress.

The cost-benefit of a joint DTIC-DAITC AI laboratory, with the sharing of its resources, makes this option an obvious recommendation.

3. SYSTEM CONFIGURATIONS

Developing and incorporating AI applications present several problems that, for the most part, effect the core system. Incorporating AI programs into DGIS means creating lengthy programs that use extensive memory, and because of that, they will take time to initialize when needed. The memory storage needs of AI programs implies that the programs should be planned for and managed so as not to siphon away the memory resources needed by the users on the core system.

It is recommended that AI development activity take place off the core system, in workstation mode. This way, the core system memory is not reduced for DGIS users because of development programming. A second recommendation is that AI-based programs made available for production use also reside off the core system. This may be accomplished through program storage on the processor workstation where the development is done, to serve in consultant mode to the user; or better, a separate machine that serves as a backend to the core system, where programs may be stored and called upon as needed by DGIS users.

The following figures show these two possible, basic system configurations for AI applications development and implementation.

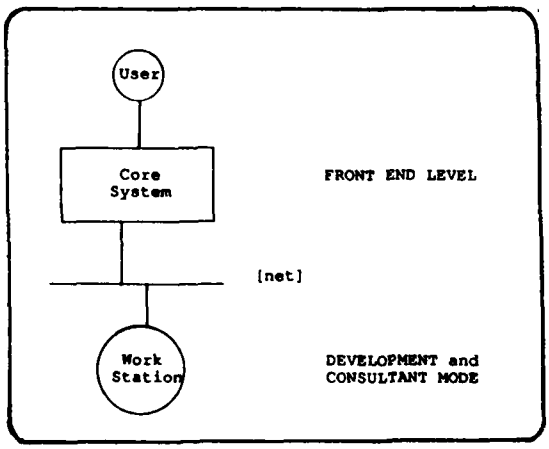


Figure: Workstation Consultant Mode

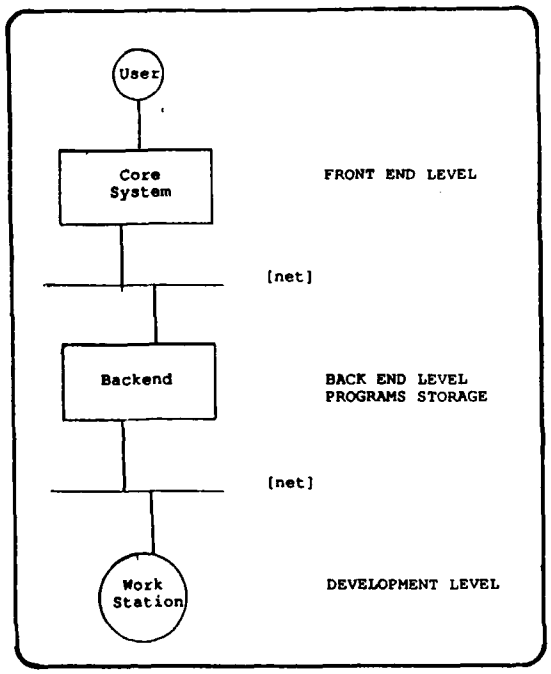


Figure: Backend Storage Mode

It is recommended that the Backend Storage Mode be employed, for optimum management not only of AI implemented programs, but also for other programs which may have lesser use, but can be initialized in a facile manner, without decreasing the memory resources of the core system to the detriment of the other users.

IV. PROJECTED COSTS

The following start-up cost projection is based on a configuration model of four AI experts and appropriate hardware/software. Conjectured costs are based on information collected during FY 87 Qtrs 1 through 3.

NOTE: The hardware and software shown in the cost models below are by no means an indication nor a commitment on the part of DTIC to procure them. They are only used as examples of costs involved in establishing an AI environment.

1. One-Time Start-Up Cost Model -

Expertise:	4 Masters/PhD level Fellows @ $125K pa:	$500K
Hardware:	3 Symbolics Processors @ $80K:	$240K
	1 TI Workstation @ $75	$ 75K
Software:	LISP, PROLOG, KEE, S-1, ART	$200K
ONE TIME START-UP COSTS TOTAL		$1015K

2. Fiscal Year Incremental Start-Up Cost Model -

	FY 1	FY 2	FY 3
Expertise	2 Fellows @$125K/ $250K	1 Fellow $130K	1 Fellow $135K
Hardware	1 Symbolics $80K 1 TI Wkstation $75K	1 Symbolics $85K	1 Symbolics $90K
Software	1 LISP 1 PROLOG 1 ART 1 KEE 1 S-1 $200K		
TOTALS	$605K	$215K	$225K

INCREMENTAL START-UP COSTS TOTAL: $1045K

V. SUMMARY

1. PLAN AND MISSION

A DTIC goal is to increase database services directly to endusers. DTIC's endusers are the research and development (R&D) managers, engineers, and scientists. These users depend on accurate and up-to-date information to fulfill their mission assignments. Proliferation of microcomputers makes database access feasible for the DoD enduser.

A major obstacle for enduser database utilization is knowing what and how to search. These users need a system that will allow them to identify, access, and query databases using natural language dialogues. Artificial intelligence technology applied to information retrieval systems provides potentials to fulfill these user requirements. The establishment of an Artificial Intelligence environment will be the mechanism for the development and incorporation of AI applications in the DoD Gateway Information System and wherever else needed.

The AI activity needs both expertise and materiel to function. The AI expertise required should at the minimum be at a Master's level with extensive C/UNIX and AI programming backgrounds. A PhD. level expert is to have this level of studies in AI applications relevant to the needs of the DGIS programs. The hardware and software requirements shown above will allow development of programs incorporating the use of expert systems and use of natural language as a means of providing the needed information querying capabilities to endusers. AI activities, after they have developed, could further lead into exploring other DoD enduser needs such as technical information networking and transfer, intelligence programs, and industrial/manufacturing programs assessment.

2. BENEFITS

The application of AI to information retrieval will allow DoD R&D managers, engineers, and scientists to overcome the barriers of user-system communications. They will be able to more easily access and utilize the information they need. The availability of this information will make our endusers more efficient and help them avoid the high costs of duplicating or ignoring pertinent work. The incorporation of AI applications in DGIS will additionally make the human-machine interface on DGIS more human-like in its functions and responses, and tolerant of human frailties that are caused by the complexities of the human mind.

REFERENCES

1 Kuhn, Allan D. & Tran, Duc T. Toward an Artificial Intelligence Environment for DTIC: Staffing Qualification Criteria for AI Application Development. Defense Technical Information Center, February 1987, AD-181 100.

2 Cotter, Gladys A. The DoD Gateway Information System. Defense Technical Information Center, October 1985, AD -A161 701.

3 Cotter, Gladys A. The DoD Gateway Information System: Prototype Experience. Defense Technical Information Center, April 1986, AD-A166 200.

4 Kuhn, Allan D. Artificial Intelligence Developments Re: DoD Gateway Information System ((DGIS) & Defense Applied Information Technology Center (DAITC). Defense Technical Information Center, February 1987, AD-A181 101.

END

7-87

DTIC